



## King's Research Portal

DOI:

[10.1111/rssc.12170](https://doi.org/10.1111/rssc.12170)

*Document Version*

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Parker, B. M., Gilmour, S. G., & Schormans, J. A. (2017). Optimal design of experiments on connected units with application to social networks. *APPLIED STATISTICS*, 66(3), 455-480. <https://doi.org/10.1111/rssc.12170>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Optimal Design of Experiments on Connected Units with Application to Social Networks

Ben M. Parker<sup>\*1</sup>, Steven G. Gilmour<sup>1</sup> and John Schormans<sup>2</sup>

<sup>1</sup>Southampton Statistical Sciences Research Institute, University of Southampton, UK

<sup>2</sup>School of Electronic Engineering and Computer Sciences, Queen Mary University of London, UK

June 1, 2016

## Abstract

When experiments are performed on social networks, it is difficult to justify the usual assumption of treatment-unit additivity, due to the connections between actors in the network. We investigate how connections between experimental units affect the design of experiments on those experimental units. Specifically, where we have unstructured treatments, whose effects propagate according to a linear network effects model which we introduce, we show that optimal designs are no longer necessarily balanced; we further demonstrate how experiments which do not take a network effect into account can lead to much higher variance than necessary and/or a large bias. We show the use of this methodology in a very wide range of experiments in agricultural trials, and crossover trials, as well as experiments on connected individuals in a social network.

**Key words:** Design of Experiments; Social Networks; Optimal Design.

## 1 Introduction

We investigate how the structure of relationships between experimental units affects the design of experiments on these experimental units. We consider examples where the treatment applied to a single experimental unit has an effect on further experimental units which have some relationship to the one given the treatment. Previous research has focused on agricultural trials, where the treatment applied to an experimental unit (a plot) also has an effect on neighbouring experimental units within a block, and on cross-over experiments where a treatment continues to have an effect on the subject when measured at subsequent time periods. That research generally considers a class of experiments in which the subjects are related according to some regular pattern, and here we remove any restrictions on the regularity of the relationships.

Our primary application is in investigating experiments on social networks. These networks are prevalent in virtual online communities, such as Facebook or Twitter, but the methodology

---

<sup>\*</sup>B.M.Parker@soton.ac.uk

also encompasses networks in real world communities, such as friendship, familial relationship, or geographical networks. We consider the case where an experiment is to be performed on a known social network and it is expected that the treatments applied will affect not only the nodes of the network to which they are applied, but also all those nodes which are attached to them. This makes it difficult to make the usual assumption in randomized experiments that treatment effects are additive, i.e. the treatment effects have the same distribution no matter which experimental unit they are applied to. This idea is familiar from cross-over trials with carry-over effects and agricultural field experiments with neighbour effects, but the neighbour structures in these applications are usually very regular. Social networks typically have irregular, unbalanced patterns of connections and dealing with this requires some new design methodology. We also adapt this research on social networks to non-standard examples in finding the optimal design of experiments for agricultural trials and crossover trials.

### 1.1 Motivating example

In online marketing experiments different advertisements, or different versions of the same advertisement, are investigated to find out which are most effective. For example, an advertisement for a product may describe it as either “half price” (1) or “50% off” (2), or some other alternative phrasings, which we define as treatments in the usual terminology. In other cases the treatments might have a factorial structure, though we will not deal with that case here.

For our motivating example, we assume we have a network of 10000 customers, and want to assess the effectiveness of sending 4 different adverts to a subset of 100 customers (before we send the best advertisement to the rest of the customers). An advertisement is deemed effective if more customers click on a hyperlink in a message. The advertisements will be released to a test panel of subjects, with each subject being allocated a single treatment. The response of these subjects to the advert will be measured, by measuring positive the number of times that hyperlinks are clicked in the advertisement. (Alternatively we might choose to measure a longer term response, such as a measure of changes in the behaviour of the subjects.)

The standard randomised experiment requires that the responses from subjects are independent; in reality, especially with online marketing, advertising to one person will often mean that the advertisement is passed to a second person. Indeed, this viral effect is sometimes exactly what marketers wish to encourage: advertisers pay for distribution to an original population, with popular adverts distributed for free to those in the original target’s social network.

We set out in this paper to investigate how experimental design might take into account the structure of a social network, to get the optimal information about which adverts are best when experimental units are not independent. We see that these social networks readily generalise to enable us to consider a framework which applies to many situations for designing experiments on connected experimental units.

### 1.2 Outline of this paper

In Section 2 we set out the model considered in this paper, and review previous work in designing experiments on connected experimental units. We introduce terminology and the

methodology for designing experiments on this model in Section 3, before showing the effect of misspecification of the model in Section 4. The importance of the network structure is discussed in Section 5, and algorithms for finding an optimal design in Section 6. We present a collection of examples in Section 7, before we conclude briefly in Section 8.

## 2 Models Considered and Previous Work

We consider networks of subjects; a network  $G = (N, E)$  is an undirected graph, a collection of nodes  $N$  and edges  $E \subseteq (N \times N)$  where the nodes represent experimental units (typically subjects) on which we apply some treatment. The edges represent relationships between the subjects.

We assume that if a relationship exists between two subjects, the response of each subject is dependent on the treatment applied to the other according to a linear network effects model which we specify below.

We have  $|N| = n$  subjects, and  $m$  treatments. The relationship between subjects is specified by the  $n \times n$  adjacency matrix  $A$  where  $A_{ik} = 1$  if and only if  $i$  and  $k$  are related and  $A_{ik} = 0$  otherwise. By convention,  $A_{ii} = 0$ . For ease of exposition in this paper, we assume initially that links are non-directional, such that  $A_{ik} = A_{ki}$ , i.e.  $A$  is a symmetric matrix.

We assume there is a “subject effect” on the response from subject  $i$  of  $\tau_j$  when treatment  $j$  is given to that subject, and a “network effect” of  $\gamma_l$  if treatment  $l$  is given to a connected subject  $k$  (if  $A_{ik} = 1$ ). We assume that each subject gets exactly one treatment.

### 2.1 Linear network effects model

We measure the response  $Y_i$  for each of our subjects. Let  $t(i)$  be the treatment applied to subject  $i$ ; then our response is modelled as

$$Y_i = \mu + \tau_{t(i)} + \sum_{k=1}^n A_{ik} \gamma_{t(k)} + \epsilon_i. \quad (1)$$

We assume that errors  $\epsilon_i$  are independent and identically distributed with mean 0 and constant variance  $\sigma^2$ . We assume we wish to estimate the subject and/or network effects, or some contrast of them.

We let  $\mathbf{u}_j$  be the vector with  $i$ -th element equal to 1 when treatment  $j$  is applied to subject  $i$ , and otherwise 0. (See section 2.7 of Bailey [2008], for example.) In matrix form,

$$\mathbb{E}(\mathbf{Y}) = (\mathbf{1} \quad \mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_m \quad A\mathbf{u}_1 \quad A\mathbf{u}_2 \quad \dots \quad A\mathbf{u}_m) \begin{pmatrix} \mu \\ \boldsymbol{\tau} \\ \boldsymbol{\gamma} \end{pmatrix} = F\boldsymbol{\beta}, \quad (2)$$

where  $F$  is our design matrix, and our vector of parameters is

$$\boldsymbol{\beta} = (\mu \quad \boldsymbol{\tau}^T \quad \boldsymbol{\gamma}^T)^T = (\mu \quad \tau_1 \dots \tau_m \quad \gamma_1 \dots \gamma_m)^T. \quad (3)$$

We let the design space, the set of possible experiments we may perform, be

$$\mathcal{X} = \{[t(1), \dots, t(n)] | t(i) \in \{1, \dots, m\}, i = 1, \dots, n\},$$

such that we give one and only one treatment to each subject. We calculate the Fisher information matrix as  $I = F^T F$ .

## 2.2 Previous work in designing experiments for connected experimental units

Previous work in experimental design for connected units has been limited, and focused on agricultural trials, where consideration is occasionally given to how the response of one plot is affected by the treatment given to neighbouring plots, as well as the treatment given to the plot itself (which is often called the “direct treatment effect”, though we refer to it in this paper as the subject effect).

Besag and Kempton [1986] review the analysis of field experiments where responses in a plot are affected by the treatments applied to neighbouring plots.

Druilhet [1999] considers designing experiments in linear blocks, with block effects and with left and right neighbour effects only in the direction of the blocks, in the context of both agricultural trials, and crossover trials. The model used is  $\mathbb{E}(\mathbf{Y}) = b\boldsymbol{\beta} + T\boldsymbol{\tau} + L\boldsymbol{\lambda} + R\boldsymbol{\rho}$  where  $\boldsymbol{\beta}$  are block effects,  $\boldsymbol{\tau}$  are treatment effects,  $\boldsymbol{\lambda}$  are left-neighbour effects,  $\boldsymbol{\rho}$  (sometimes assumed to be zero for uni-directional effects) are right-neighbour effects, and  $B, T, L$ , and  $R$ , are incidence matrices for block, treatment, and neighbour effects. He discusses circular neighbour-balanced designs, where the number of times treatment  $i$  is to the left of treatment  $j$  is constant, for  $i \neq j$ , and shows these are optimal for the estimation of treatment effects under certain conditions.

These two previous papers both assume that there is an edge plot, or guard plot, such that each plot of interest (the interior plots) has two neighbours. Kunert and Martin [2000] relax this assumption, but still assume that there are one-dimensional arrangements of plots in each block (i.e. that a field trial has a left and right neighbour effect, but no up or down effect). They present criteria for designs to be optimal and, for certain examples with block size up to four, demonstrate that there are designs which fulfill these criteria, but do not show how to construct optimal designs. Kunert and Mersmann [2011] extend this to larger block sizes. Azaïs and Druilhet [1997] considers minimising bias in this model, and shows that neighbour-balanced (randomized) designs minimize bias.

The models considered in Druilhet [1999], Kunert and Martin [2000], and Kunert and Mersmann [2011] are, with left and right neighbour effects assumed to be equal and neglecting the block effects, a special case of the Linear Network effects model (equation (1)) we consider in this paper.

Pearce [1957] assumes a treatment affects the plot it is applied to (a “local effect”), and also other plots in the same block (a “remote effect”). He allows interaction between these effects, i.e the remote effect of a treatment on a plot depends on the treatment applied locally to that plot. He shows orthogonal designs no longer guarantee all effects are estimable, and proposes balanced incomplete block designs with restrictions of the number of treatments to maintain estimability.

In designs for cross-over trials, subjects receive sequences of treatments, and a carry-over effect is often assumed to the next time period. We show in this paper how to represent these cross-over designs as a generalisation of our model; particular examples have been studied. e.g. Hedayat and Zhao [1990] considers two period cross-over designs, and shows in particular that optimal designs for estimating the carry-over effects are not always balanced.

These temporal effects can be thought of in the same framework as treatments affecting spatially neighbouring subjects in agricultural experiments. See, for example, Jones and Kenward [2003].

There has been little work discussing designing experiments on subjects that are connected by a social network. In McConnell, Sinclair, and Green [2010], the stable unit treatment value assumption (Rubin [1978]) is discussed, and the assumption is questioned by detecting whether there is a “spillover effect” in an experiment on voter mobilization where voters are sent mailings before an election. Here a multilevel model is applied and household and neighbourhood effects are detected, where the household and neighbourhood effects have a regular pattern within nested blocks.

### 2.3 Previous work on inference for social networks

A review of modelling social influence was performed by Leenders [2002]; the problem considered is to infer relationships between actors (nodes) in a social network by measuring some response on those networks. This paper considers two different types of models

$$\mathbf{y} = \rho W \mathbf{y} + X \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 I). \quad (4)$$

Here  $\rho$  is a scalar, and  $W$  a matrix representing the strength of relationships between actors.

In this model an actor in a social network forms his response from the (weighted) responses of actors connected to him. This is the network effects model of Doreian et al. [1989], sometimes called a spatial autoregressive model. In a social network, this corresponds to an actor basing his opinion (response) on his own innate response, which is modified by the actual response of his peers.

The second class of model considered is

$$\mathbf{y} = X \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} = \rho W \boldsymbol{\epsilon} + \boldsymbol{\nu}, \quad \boldsymbol{\nu} \sim N(0, \sigma^2 I). \quad (5)$$

Here the response of an actor in a social network depends on the *change* in response of actors connected to him. This network is sometimes called a spatial moving average model. In social networks, these correspond to an actor altering his opinion (response) due to a change in opinion of his peers; as we demonstrate below, this paradigm seems to be valuable in physical experiments where changes in an experimental unit may also propagate somewhat to its neighbours because of some spillover or neighbourhood effect caused by applying a treatment.

These two form a wide class of models, and mathematical interest is found in calculating the strength of actors’ relationships (the weight matrix  $W$ ). The non-network effects  $\boldsymbol{\beta}$  can be inferred using maximum likelihood and appropriate hypothesis testing carried out to investigate whether a particular network structure exists. The review paper, Leenders [2002], emphasizes the large effect that  $W$  has on any conclusions, and the importance of correct model choice.

## 2.4 Experiments covered in our model

To begin with the simplest case, we focus on the Linear Network effects model(1). Although a relatively simple model, this allows us clearly to focus on demonstrating the optimal design of experiments on networks, a novel methodology. In the methodology described in this paper, and unlike in much previous research in designing experiments, we do not impose any restrictions or symmetries in the relationships between subjects, such that any two subjects may be assumed related or not. We present a linear network effects model for unstructured treatments within this framework, and discuss how we can find optimal designs for experiments in this case.

This setting admits many experiments, for example in our motivating example the nodes are subjects in a marketing experiment, and the links are a social relationship, such as being “friends” on Facebook or some other social network. The treatments are different adverts, and the responses the number of hyperlinks clicked by a subject. If a subject is exposed to marketing, their friend may be affected by this marketing and the friend’s response may be altered (positively or negatively). If a subject is given some information about a product in a marketing campaign, he may tell a friend about it, which may affect the number of times the friend clicks a hyperlink.

In another example, the nodes may be subjects in a public health experiment. The links may be some friendship relationship, some geographical relationship, or some familial relationship. The treatments may be different messages given in the campaign (e.g. “Eat five fruit or vegetables a day”) and the response the change in fruits or vegetables eaten after the campaign. A member of a family receives a treatment and, whether or not they alter their own behaviour, it may bring about a change in their relation’s response; here the nodes are family members and the edges represent the relation “eat together”. See section 7 for further examples of the method in agriculture, medicine, and social networks.

We do not consider spatial autoregressive models here. For example if a child copies his parent’s behaviour in response to a public health campaign, this is described by a spatial autoregressive model as the parent’s response (and not the parent’s treatment) affects the child’s response.

## 3 Designing Experiments

Writing  $n_i$  as the number of subjects given treatment  $i$ ,  $n_{ij}$  as the number of connections between subjects given treatment  $i$  and those given treatment  $j$ , and  $n_{ij}^{(2)}$  as the number of connections between subjects given treatment  $i$  and subjects given treatment  $j$  with a path of length 2, including trivial paths that use the same edge twice, then we can write the Fisher information matrix as

$$I = F^T F = \begin{pmatrix} n & \mathbf{1}^T X & \mathbf{1}^T A X \\ X^T \mathbf{1} & X^T X & X^T A X \\ X^T A \mathbf{1} & X^T A X & X^T A^2 X \end{pmatrix} \quad (6)$$

where we recall that as  $A$  is symmetric,  $A^T = A$ . This form is useful as it separates the information due to the subject effects and the network effects (e.g. the inverse of the top-

left submatrix is the covariance matrix for the least squares estimate of the subject effects). Expanding this gives

$$I = \begin{pmatrix} n & n_1 & n_2 & \dots & n_m & \sum_i n_{i1} & \sum_i n_{i2} & \dots & \sum_i n_{im} \\ n_1 & n_1 & 0 & \dots & 0 & n_{11} & n_{12} & \dots & n_{1m} \\ n_2 & 0 & n_2 & \dots & 0 & n_{21} & n_{22} & \dots & n_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ n_m & 0 & 0 & \dots & n_m & n_{m,1} & n_{m,2} & \dots & n_{m,m} \\ \sum_i n_{i1} & n_{11} & n_{12} & \dots & n_{1,m} & n_{11}^{(2)} & n_{12}^{(2)} & \dots & n_{1m}^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sum_i n_{im} & n_{m1} & n_{m2} & \dots & n_{m,m} & n_{m1}^{(2)} & n_{m2}^{(2)} & \dots & n_{mm}^{(2)} \end{pmatrix}. \quad (7)$$

As this matrix represents the information we can gain about the experiment performed with design matrix  $F$ , we wish this matrix in some senses to be as “large as possible”. Loosely speaking, we wish  $\det(A)$  to be large (this is the D-criterion for optimality, although in this paper we use the criteria described in Section 3.1 below). Designs seem to have better optimality criterion function values when they have

- an approximately balanced number of subjects given each treatment;
- a large number of occasions where the same treatment is applied to two neighbours of a subject;
- and a small number of occasions where the same treatment is applied to neighbours.

Of course, these features can conflict, but this gives an idea as to which properties are important in a good design. We find the optimal design by maximising some optimality function of this matrix (see section below) subject to some constraints (such as  $\sum_{i=1}^m n_i = n$ , and other constraints particular to the network). We might consider a constrained optimisation to find the optimal design. However, in all except trivial networks, the function to be optimised is complicated (even for  $m = 2$  it is a quartic) and, computationally, a random or exhaustive search is preferable in order to find the optimal design.

### 3.1 Optimality Criteria

We can find the optimal design for different functions of the covariance matrix of the parameter estimates. For any possible design  $X \in \mathcal{X}$  we could reasonably seek to minimise the average variance

1. of all pairwise differences of treatment effects,

$$\frac{2}{m(m-1)} \sum_{j=1}^{m-1} \sum_{l=j+1}^m \text{Var}(\widehat{\tau_j - \tau_l}).$$

This is defined to be A-optimality for estimating the differences in the treatment effects. We call this criterion function  $\phi_1$ .



2. of all pairwise differences of network effects,  $\gamma_j$ . This is defined to be A-optimality for estimating the differences in the network effects. We call this criterion function  $\phi_2$ .

Let  $\mathbf{c}(a, b)$  be a length  $2m + 1$  vector which is a vector of zeroes, except the  $a$ -th element is 1 and  $b$ -th element is  $-1$ . Recall that  $I$  is a  $(2m + 1) \times (2m + 1)$  matrix, the first row/column corresponding to the mean  $\mu$ , the next  $m$  elements to the subject effects  $\boldsymbol{\tau}$ , and the final  $m$  elements to the network effects  $\boldsymbol{\gamma}$ . Then the optimality criteria above are equivalent to minimizing

1.  $\phi_1 = \sum_{j=2}^m \sum_{l=j+1}^{m+1} \mathbf{c}^T(j, l) I^{-1} \mathbf{c}(j, l)$ .
2.  $\phi_2 = \sum_{j=m+2}^{2m} \sum_{l=j+1}^{2m+1} \mathbf{c}^T(j, l) I^{-1} \mathbf{c}(j, l)$ .

Other criteria may be appropriate depending on the aims of the experiment (e.g. a weighted sum of criteria 1 and 2). We call the values of these optimised functions at the optimal design the optimal function values, and represent them as  $\phi_1^* = \max_{\mathcal{X}} \phi_1$ , etc.

We consider primarily the first criterion in the examples presented in this paper. Note that if we have  $n_j$  subjects given treatment  $j$ , we could (ignoring the network effect) estimate  $\tau_j$  with variance  $\sigma^2/n_j$ . Thus we can estimate each difference  $\tau_j - \tau_l$  with variance  $\sigma^2/n_j + \sigma^2/n_l$ . If we have an equi-replicate design such that  $n_j = n/m \quad \forall i$ , then the minimum possible variance for the subject effects is  $2\sigma^2/n_j = 2\sigma^2 m/n$ .

### 3.2 Estimability

In practical situations, we are interested in the (variance of) the difference between treatments: whether advert 1 is more effective than advert 2, for our motivating example. As is common in this situation, we will always have to impose some additional constraint on our parameters  $\beta = (\mu, \tau, \gamma)^T$  to ensure they can be estimated uniquely. Without loss of generality, for the rest of this paper we have assumed that  $\tau_m = 0$ .

Even with this assumption, not all networks allow all effects to be estimated; consider the network where each node is connected to every other node. Then if subject  $i$  is given treatment  $j$  the response is

$$\begin{aligned} \mathbb{E}(Y_i | t(i) = 1) &= \mu + \tau_1 + (n_1 - 1)\gamma_1 + n_2\gamma_2, \\ \mathbb{E}(Y_i | t(i) = 2) &= \mu + \tau_2 + n_1\gamma_1 + (n_2 - 1)\gamma_2. \end{aligned}$$

As  $\tau_2 = 0$ , we have only two different equations with which to estimate four unknowns (the mean  $\mu$ , subject effect  $\tau_1$  and network effects  $\gamma_1$  and  $\gamma_2$ ), and thus these are not estimable. Less extreme examples exist, with a number of edges less than the maximum possible.

In general for a linear model,  $L\beta$  is estimable if and only if there is a  $K$  such that  $L = KF$ . A special case for all elements of  $\beta$  to be estimable, where  $L$  is the identity matrix, implies that  $K$  must be a generalised inverse  $F^-$  of  $F$ . A sufficient condition for this would be that  $(F^T F)^{-1} = I^{-1}$  exists such that  $F^- = (F^T F)^{-1} F^T$ , i.e. that  $I$  can be inverted, hence  $\det(I) \neq 0$ .

From the definition of  $I$  in equation (7), it is not clear what properties of a given network structure combined with a design lead to estimability; we can show trivially that  $n_i$  must be

greater than zero (that all treatments must be given to at least one subject) otherwise a row and column of  $I$  will be zero, but further general statements applicable to all networks may not be possible. All the networks described in this paper are estimable with the sole assumption that  $\tau_m = 0$ , but further assumptions may have to be made for particular networks.

### 3.3 Example 1

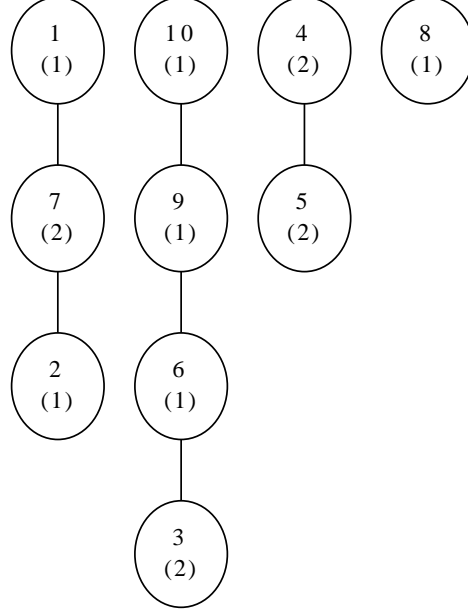


Figure 1: An example network; node numbers are shown above with the optimal design for estimating  $\tau_1$  for two treatments shown below as the number in brackets

Let us suppose we have a network of 10 subjects, where some subjects are related to other subjects as shown in Figure 1. For example, subject 7 is connected to subjects 1 and 2 only, but subject 8 is not connected to any other subject.

By exhaustive search over all possible designs, we find that the optimal design for the  $m = 2$  treatment case is  $\{1, 1, 2, 2, 2, 1, 2, 1, 1, 1\}$  for optimality in estimating  $\tau_1$  via criterion  $\phi_1$  (i.e. we give treatment 1 to subjects 1, 2, 6, 8, 9, and 10 and treatment 2 to the other subjects). This design is that shown on Figure 1. The design  $\{2, 2, 1, 1, 1, 1, 2, 2, 2, 1\}$  is A-optimal for estimating the difference in the network effects  $\gamma_1 - \gamma_2$  via criterion  $\phi_2$ .

The optimal designs are different depending on what we wish to measure, with the result that the experimenter must think carefully about their experimental aims and the shape of the network before starting their experiment. This is a standard feature of optimal design.

The first design is unbalanced; i.e treatments 1 and 2 are not applied to an equal number of subjects. This is an unusual property in optimal design.

### 3.4 Example 2

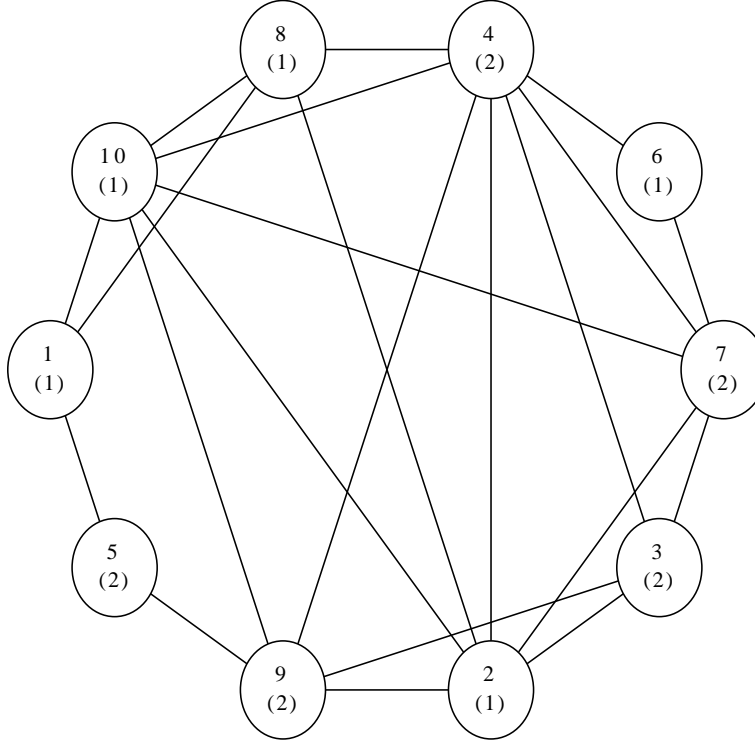


Figure 2: A second example network

We demonstrate another connectivity graph in Figure 2. This network again has 10 subjects connected by a friendship relationship, but the subjects are more likely to be connected to other subjects than in the previous network. Design  $\{1, 1, 2, 2, 2, 1, 2, 1, 2, 1\}$  is A-optimal for estimating  $\tau_1$ . This design is that shown on Figure 2, with  $\phi_1^* = 0.406$ . Design  $\{2, 1, 1, 1, 1, 1, 2, 2, 1\}$  is optimal for estimating the difference in network effects  $\gamma_2 - \gamma_1$ , with  $\phi_2^* = 1.227$ .

Again, the designs which are optimal for  $\phi_1$  and  $\phi_2$  are different. However for this network the optimal design for  $\phi_1$  in estimating the subject effects is balanced, such that 5 subjects are given each treatment, but that for estimating the network effects ( $\phi_2$ ) is not.

## 4 Misspecification of model: Efficiencies and biases

Let us now suppose we perform an experiment on subjects where we wrongly assume at the design stage that there is no network effect, although the network effect does in fact act according to our model described above in equation (1).

For  $n = 12$  subjects and both  $m = 2$  and  $m = 3$  treatments, we generate 10 Erdős-Renyi random networks where two nodes are connected with fixed probability of either 0.2, 0.4, 0.6, or 0.8, and calculate the efficiencies of all possible balanced designs (there are  $\frac{n!}{((n/m)!)^m}$  of these

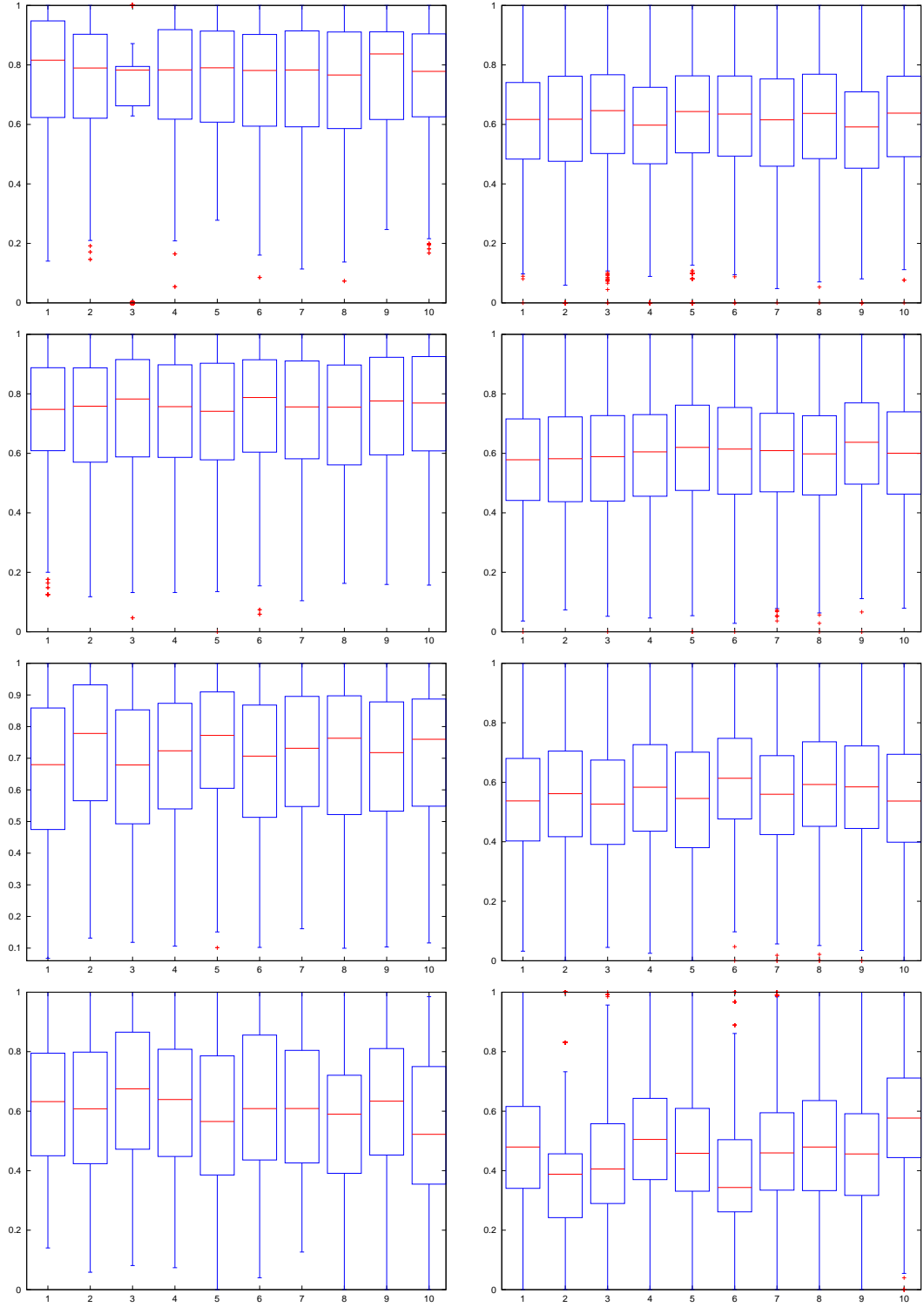


Figure 3: Efficiencies of balanced designs ignoring network effects. Efficiency (y-axis) for 10 random networks of 12 subjects and (left) 2 and (right) 3 treatments and (from top to bottom) probability of nodes connected 0.2, 0.4, 0.6, 0.8

as  $m$  divides  $n$ , so 924 for  $m = 2$  and 6930 for  $m = 3$ ), compared to the optimal design for that random network if there is a network effect. The results are shown as Figure 3.

We see that, for instance, in the top left ( $n = 12, m = 2, p = 0.2$ ) case the median efficiency of the set of balanced designs is typically around 0.8. For all networks we can find a balanced design with very bad efficiency (network 3 has a balanced design with efficiency of 0, where all subject effects are not estimable, and also has no balanced design with 100% efficiency, meaning that the optimal design is unbalanced). The range and interquartile range of efficiencies are always large; we conclude that if there is a network effect, it is vital to include it for an efficient design.

We note that as the probability of nodes being connected increases, the median efficiency decreases; also the efficiencies of the balanced design for  $m = 3$  treatments are less than those for  $m = 2$ .

#### 4.1 Example 1 revisited

Consider example 1 (see Figure 1), where  $m = 2$ . If we believed there is no network effect in the model, we would conclude that balanced designs with 5 subjects chosen at random given each of treatments 1 and 2 were optimal. We might chose an arbitrary balanced design  $\{1, 1, 1, 1, 1, 2, 2, 2, 2, 2\}$ . The design for estimating the subject effects using model (1) with this design has  $\phi_1 = 0.431$ . We saw that the optimal design had value  $\phi_1^* = 0.418$ . The efficiency of this arbitrary balanced design is thus  $0.418/0.431 = 97.0\%$ . The average variance of estimates for  $\tau_1$  is slightly bigger than with the optimal design when the network effect is accounted for.

#### 4.2 Bias of design

Now let us consider the bias of the design. If we know the adjacency matrix, and assume that there is a network effect  $\gamma \neq 0$ , such that the true model is as specified in equation (2), we can find the optimal design matrix  $X_C$  by exhaustive search as above and hence form the extended design matrix  $F_C = (\mathbf{1} \ X^* \ AX)$ , where  $X^*$  is the design matrix  $X$  without its last column to impose the constraint  $\tau_m = 0$ . From the Gauss-Markov Theorem, we find a best linear unbiased estimator for  $\beta$  under these assumptions as  $\hat{\beta}_C = (F_C^T F_C)^{-1} F_C^T Y$ .

If we perform an experiment where we wrongly assume there are no network effects, then our model is  $\mathbf{Y} = F_R(\mu \ \boldsymbol{\tau}^*)^T + \boldsymbol{\epsilon}$ ,  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 J_n)$ , with  $J_n$  being the  $n \times n$  identity matrix; then the least squares estimator of the mean, subject, and network effects  $\beta$  is  $\hat{\beta}_R = [(F_R^T F_R)^{-1} F_R^T \mathbf{Y} \ \mathbf{0}_m]^T$ , where the extended design matrix for any design which we would believe to be optimal for estimating the subject effects with no network effect can be found as  $F_R = (\mathbf{1} \ X_R^*)$ . Here  $X_R^*$  is defined as  $X_R$  with the last column deleted to impose the constraint  $\tau_m = 0$ . The biases in our estimators, under the assumption there is a network effect and we do not account for it, is  $\mathbb{E}(\hat{\beta}_R) - \beta = \mathbb{E}(\hat{\beta}_R) - \mathbb{E}(\hat{\beta}_C) = \mathbb{E}(\hat{\beta}_R - \hat{\beta}_C)$  as  $\hat{\beta}_C$  is

an unbiased estimator. We can thus find the bias as

$$\begin{aligned}\mathbb{E}(\widehat{\beta}_R - \widehat{\beta}_C) &= \mathbb{E} \left[ \begin{pmatrix} (F_R^T F_R)^{-1} F_R^T \\ \mathbf{0}_{m \times n} \end{pmatrix} \mathbf{Y} - (F_C^T F_C)^{-1} F_C^T \mathbf{Y} \right] \\ &= \left[ \begin{pmatrix} (F_R^T F_R)^{-1} F_R^T \\ \mathbf{0}_{m \times n} \end{pmatrix} - (F_C^T F_C)^{-1} F_C^T \right] \mathbb{E}[\mathbf{Y}] \\ &= \left[ \begin{pmatrix} (F_R^T F_R)^{-1} F_R^T \\ \mathbf{0}_{m \times n} \end{pmatrix} - (F_C^T F_C)^{-1} F_C^T \right] F_C \beta,\end{aligned}$$

which is not in general zero, and neither are the elements corresponding to just the subject effects (elements 2 to  $m$  of this expectation vector). In general, the bias of any contrast of a subject effect  $\tau_j - \tau_l$  will also not be zero, and thus ignoring the network effects in calculating a design will produce a biased estimator of the subject effects.

In examples 1 and 2, if we wrongly assumed there was no network effect when in fact it existed, then for the balanced design  $\{1, 1, 1, 1, 1, 2, 2, 2, 2, 2\}$  which is optimal in the case there were no network effects, we can calculate the bias in terms of unknown parameters  $\beta$  as

$$\begin{pmatrix} 0 & 0 & 0.6 & 0.8 \\ 0 & 0 & -0.2 & -0.2 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \beta, \quad \begin{pmatrix} 0 & 0 & 2.8 & 1.6 \\ 0 & 0 & -1.2 & 1.2 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \beta$$

for examples 1 and 2 respectively. We of course see that the biases in the network effects (elements 3 and 4) are equal to the negative of the effects themselves, as we have ignored them; there is no bias in the estimate of  $\tau_2$  as we have assumed this to be zero and it is not included in  $\beta$ . The biases for  $\mu$  and  $\tau_1$  are (different) linear combinations of the unknown values of the network effects. If the true value of the network effects are zero (such that  $\beta_3 = \beta_4 = 0$ ) then our design will produce unbiased estimators.

If the network effects are substantial compared to the subject effects, ignoring network effects might lead to the subject effects being estimated badly with very large bias. Thus by not taking into account a network effect in our design, we produce an experiment which has higher variance than necessary, and/or biased estimators.

## 5 Effect of network structure on optimal design

We consider how properties of the network affect the design. Ideally, we would like to find a result which links the macroscopic properties of the network to the optimal design for that network.

One way of assessing the macroscopic properties of the graph is through the spectrum of the graph. In spectral graph theory, the spectrum of the graph is the set of eigenvalues of the adjacency matrix  $A$ . The Laplacian spectrum is also often studied, i.e the eigenvalues of the Laplacian  $L$  where  $L = D - A$ . The degree matrix  $D$  specifies the number of neighbours each node has:  $D_{ii} = \sum_j A_{ij}$ ,  $D_{ij} = 0$  for  $i \neq j$ .

These eigenvalues have various interpretations and summarise some important macroscopic properties of the graph; for example, the second smallest eigenvalue of the Laplacian matrix

is the algebraic connectivity of our network. This gives a measure of how well connected the overall graph is. It is reasonable to suspect that properties of the network, such as algebraic connectivity, as expressed through the spectra, affect the optimal design for that network.

To test this hypothesis we formed seven Erdős-Renyi random graphs with  $n = 6$  nodes, where each pair of nodes is connected with equal probability 0.5. We calculated the spectrum and the Laplacian spectrum.

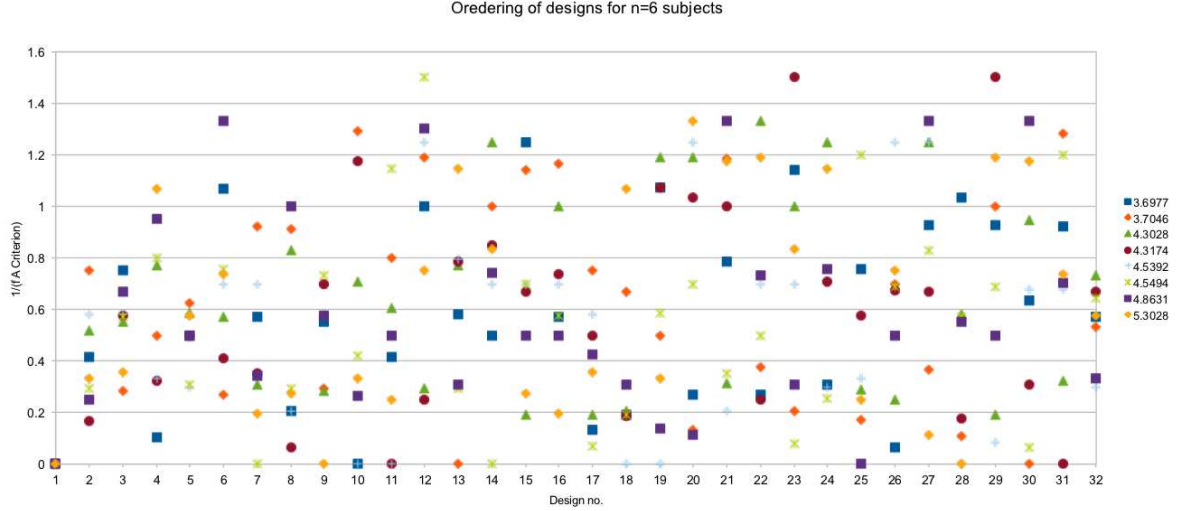


Figure 4: Algebraic connectivity for the optimal design for different networks

We plot the results for this as Figure 4. Each vertical series on the x-axis corresponds to one of the seven different networks, with the network's algebraic connectivity labelled. The y-axis is the reciprocal of  $\phi_1$  (for  $m = 2$  and  $n = 6$  the theoretical maximum is 1.5), such that a higher point corresponds to a better design. The horizontal series correspond to six (randomly chosen) possible designs. Consider, for example Design 5. It is the best of the six designs for the network with algebraic connectivity 4.3174, joint best for that with algebraic connectivity 4.3028, very bad for networks with algebraic connectivity 3.7046 and 4.5494, and mediocre for the rest.

We see that each different design may be good, bad, or indifferent for each particular network. There seems to be no fixed ordering, such that designs tend to be better for all networks. There seems to be no correlation between the quality of a design and the algebraic connectivity.

As we have seen before, for some networks we achieve no information from the experiment, so here  $\tau_1$  is not estimable (represented as those points on the x-axis). The network with algebraic connectivity 4.3174 has one such design.

We see similar patterns for other eigenvalues and Laplacian eigenvalues, but omit them due to space, and also due to the fact that their physical interpretation is less clear.

As a further example to show that macroscopic properties of a network do not determine the design, consider the two graphs pictured as Figure 5. These graphs are clearly very different,

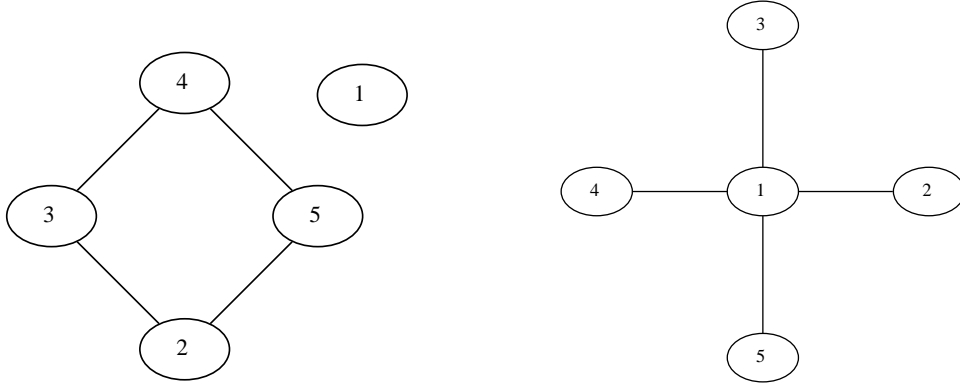


Figure 5: Two co-spectral graphs

although they have the same spectrum  $(-2, 0, 0, 0, 2)$ . However, the first graph has an optimal design with A-optimality function value 2; in fact, of the 16 distinct designs for  $n = 5$  subjects, 8 of them are optimal, and 8 are sub-optimal, and in this case lead to inestimability of the subject effects. For the second graph all designs lead to an inestimable difference in subject effects; i.e. there is no optimal design, or rather all designs are equally ineffective.

Wilson (<http://www-users.cs.york.ac.uk/~wilson/cospectral.html>) lists a number of graphs which are co-spectral in their adjacency matrix, Laplacian, as well as two other criteria (sign-less Laplacian, and normalised Laplacian). We compare the pair with upper-right triangle matrices of A

$$\begin{array}{l} 011111010101010111010000100010111001101011011 \\ \text{and } 110100101111100000011011011010100111001111010 \end{array}$$

For example, with  $m = 3$ , the  $\phi_1^*$  are 0.6263 and 0.6274 respectively. The  $\phi_2^*$  are also different (1.158 and 1.350 respectively), with different optimal designs for the two co-spectral networks. We again see that knowing the spectrum or Laplacian spectrum, does not enable us to determine which design is optimal, without also knowing the detailed structure of the network.

### 5.1 Summary of relationship between network properties and optimal design

When designing an experiment for a network, we thought initially that it was reasonable that there would be strong relationship between the large-scale properties of the network and the optimal design for that network; as many of the large scale properties of the networks are encoded in the spectrum of the graph (and e.g. Laplacian spectrum), it is surprising that two graphs which are co-spectral produce very different designs. Thus we can be sure that the large-scale properties alone of the graph do not determine the optimal design.

It may be that the large-scale properties of a graph do not determine the design absolutely, but there is an association between these properties and the design; Figure 4 shows clearly



that even the simpler properties, such as algebraic connectivity, do not have a strong effect on determining the design. Thus for any network, it is important to know the connectivity structure of the graph exactly in order to be confident of a good design;

In the case of online social networks, for some networks (e.g. Twitter) the structure of the network is publicly available, whereas for other networks, such as Facebook, user privacy prevents anyone but an administrator knowing the global connectivity structure of the network. For less well defined social networks, such as friendship networks, there may be uncertainty in a connection. Finally, the strength of connections on a network (corresponding to a non-binary matrix  $A$ ) may also strongly affect the design. Future work will include generalising this work to the case where we do not know the complete network structure, but have some information on it.

## 6 Algorithms for design

For small networks, we exhaustively consider each design within the design space  $\mathcal{X}$ , calculate its Fisher information matrix, and hence the value of the design criterion. The design which yields the best value of the design criterion function is our optimal design for that criterion. The exhaustive search is computationally expensive;  $|\mathcal{X}| = m^n$ , which becomes unmanageable for moderate network sizes  $n$  or treatments  $m$ . We can reduce the size of the search region by two methods: considering symmetry of labels in the design, and symmetries in the graph.

### 6.1 Symmetry of labels

For our criteria we are only interested in differences between treatments, the treatment effects themselves are irrelevant, and treatments are equivalent up to relabelling. For example, 1,2,2,3,1 and 1,3,3,2,1 are equivalent. We can thus reorder any design so that we only evaluate designs where the first occurrence of label  $j$  must come before the first occurrence of label  $j + 1$ .

### 6.2 Symmetry of graphs

In Figure 6, subgraphs 1 and 2 are exchangeable; i.e if we consider subdesign A on subgraph 1, and subdesign B on subgraph 2 (call this [A1,B2]), we need not also consider [A2,B1] as by symmetry this design has the same criterion value.

We can thus reduce our design space greatly if we can identify subgraphs where the designs are exchangeable. This is equivalent to finding an automorphism for our network, a relabelling or permutation of the set  $N$  such that the edges  $E$  are preserved. This is known as the Graph Automorphism Problem. Unfortunately, it is not known whether this problem is NP-complete or not (Lubiw [1981]), and no general algorithm for reducing the design space is available.

In practice, we suggest that, except in the  $m = 2$  case, it is generally quicker computationally to find graph automorphisms to reduce the design space before performing an exhaustive search.

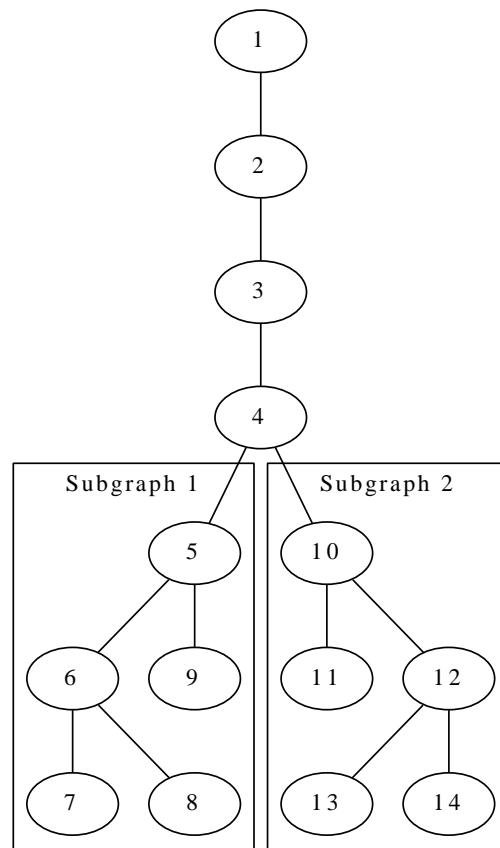


Figure 6: Demonstration of automorphism.

### 6.3 Non-exhaustive search

For large networks, exhaustive search may be difficult due to the number of possible designs ( $|\mathcal{X}| = m^n$ ). We have constructed a simple exchange algorithm to find an efficient design which may be near optimal.

1. Pick a random balanced design and evaluate the information function for that design.
2. Either
  - with probability  $p = 0.9$ : swap two randomly chosen labels
  - with probability  $1 - p = 0.1$ : pick a label randomly and relabel it with a randomly chosen label.
3. Evaluate the information function for the new design;
  - if it is smaller (better), accept the new design.
  - if it is larger (worse), accept the new design with small probability  $\delta=0.05$ .
4. Repeat from step 2 until a fixed number of iterations are completed.

Step 3 ensures that we can sometimes pick a locally worse design in the hope of eventually finding a globally optimal design. We find 10,000 iterations to be effective for the examples presented here.

## 7 Further examples

We present further examples to show how this method can find optimal designs for a wide class of experiments where the treatment applied to a subject affects other subjects.

### 7.1 Example 3: A Social Network

Let us assume we wish to design an experiment to determine the effectiveness of a new advertisement. Ten people are sent a video advertisement for a type of soft drink (treatment 1), and ten are shown no advertising (treatment 2). The amount of soft drink each person buys in the following week is then monitored. The marketing executives are interested not only in how effective the advertisement is, but also how effective it is at being conveyed to friends of those in the group which originally saw the advertisement (this is known as viral marketing), so a group of 20 people is chosen such that the connectivity structure of an online social network of the subjects is known.

We see in Figure 7 the shape of the social network for these people. The top figure for each node presents the optimal design (for  $\phi_1$ ) for estimating (with minimum variance) how much extra soft drink people buy as a direct result of having been sent the advert. If our linear network effects model is true, we find  $\phi_1^* = 0.2003$ . The bottom number in each node shows the optimal design for  $\phi_2$  for estimating the network effect (the viral effect of the advert), with  $\phi_2^* = 0.02543$ . This design is also not bad for estimating the subject effects (variance  $0.20074\sigma^2$ ).

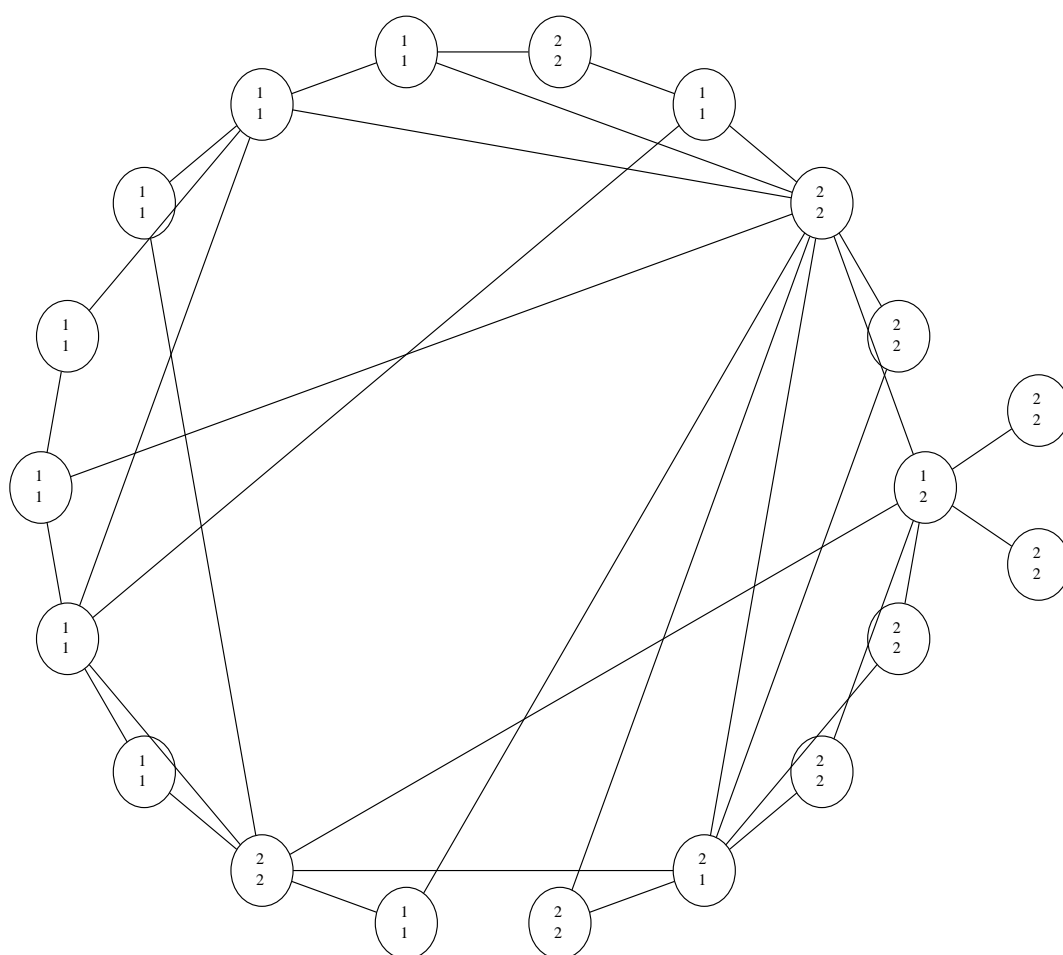


Figure 7: Optimal design for estimating the subject (top) and network (bottom) effects of an advert

## 7.2 Example 4: Block Designs for left and right treatment effects

Druilhet [1999], shows this example design is optimal for estimating  $m = 4$  treatment effects:

$$\begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & 4 \\ 1 & 4 & 3 \\ 4 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 4 \\ 3 \end{pmatrix}$$

Each number represents a treatment applied to a plot, and the treatment will affect its left and right neighbours within the block, each block being a horizontal line of plots. There is a block effect for all subjects in a block. The plots in the left and right brackets are “guard” plots, and are not measured and no response is gathered from these plots.

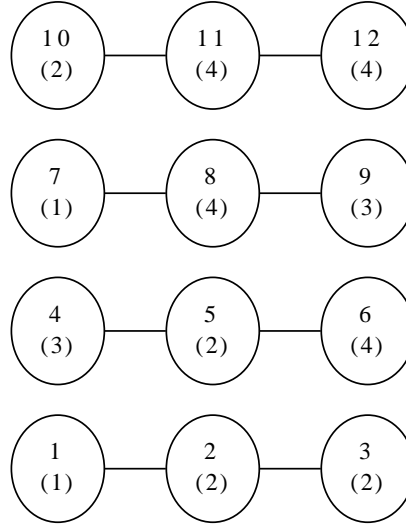


Figure 8: A blocked experiment with left and right treatment effects. Top numbers are plot labels (the nodes in the network) and the bottom numbers the optimal allocation for  $m = 4$  treatments.

We represent the plots as a network with a connectivity structure as shown in Figure 8. We also plot the  $m = 4$  optimal design for criterion 1 on that figure. Without the guard plots, it is interesting to note that the optimal design is not balanced, let alone neighbour balanced; for instance, treatment 2 appears 4 times, but treatment 1 only twice. However, in the original model, Druilhet assumes a block effect which we do not have in our present model, so a direct comparison may not be fair; with this method we can design an experiment without guard plots.

## 7.3 Example 5: Non-rectangular field trials

Most field trial experiments focus on rectangular fields. Let us assume we have an irregularly shaped field divided into plots as shown in in Table 1. An integer represents the number of

		1	2	3
		4		5
6	7	8	9	10
11	12	13	14	15

Table 1: An irregularly shaped collection of plots in a field.

a plot. We wish to perform a trial on different fertilisers, to estimate the difference between fertilizer (treatment) effects with minimum average variance. We assume that any fertiliser applied might leach to any of the 8 plots that touch the treated plot, including those only touching at corners.

Here, we can represent the connectivity structure by a network with connectivity structure as in Figure 9, shown together with the optimal design for  $\phi_1$ . This design is balanced, with all treatments equally replicated. Here,  $\phi_1^* = 0.4055$ . The minimum average variance possible with no restriction on comparisons  $2 \times \frac{\sigma^2 m}{n} = 0.4\sigma^2$ . It is interesting to note how close we can get to the variance for the unrestricted case with this unusual shape of field.

#### 7.4 Example 6: Extension to a crossover trial with planned dropouts

Usually in a crossover trial, including a drop out in the design could be tricky. Let us assume we are performing a crossover trial on four subjects (a, b, c, and d) over four periods with three treatments (labelled 1 to 3), each of which is believed to have a wash-out time of one period. Assume further that it becomes clear that participant b will not be able to take the treatment in the third period of the trial. This is represented by Table 2. We use each subject/period

		Period			
		1	2	3	4
Subject	a	1	2	3	4
	b	5	6		7
	c	8	9	10	11
	d	12	13	14	15

		Period			
		1	2	3	4
Subject	a	1	1	2	1
	b	1	3		2
	c	2	2	3	1
	d	3	3	3	2

Table 2: A crossover trial. Left: Experimental units numbered as subject/period combinations. Right: Optimal design.

combination as a node in our network, and we can modify our methodology such that links are uni-directional by relaxing the restriction that  $A$  is symmetric; our Linear Network Effects Model in Equation (1) is still valid, and we perform an exhaustive search as before to minimise  $\phi_1$ .

The network for this experiment is shown as Figure 10. The optimal design found by considering this experiment as a network is presented as Table 2. The optimal design has  $\phi_1^* = 0.4128$ . Note that it is quite different in form from usual cross-over designs.

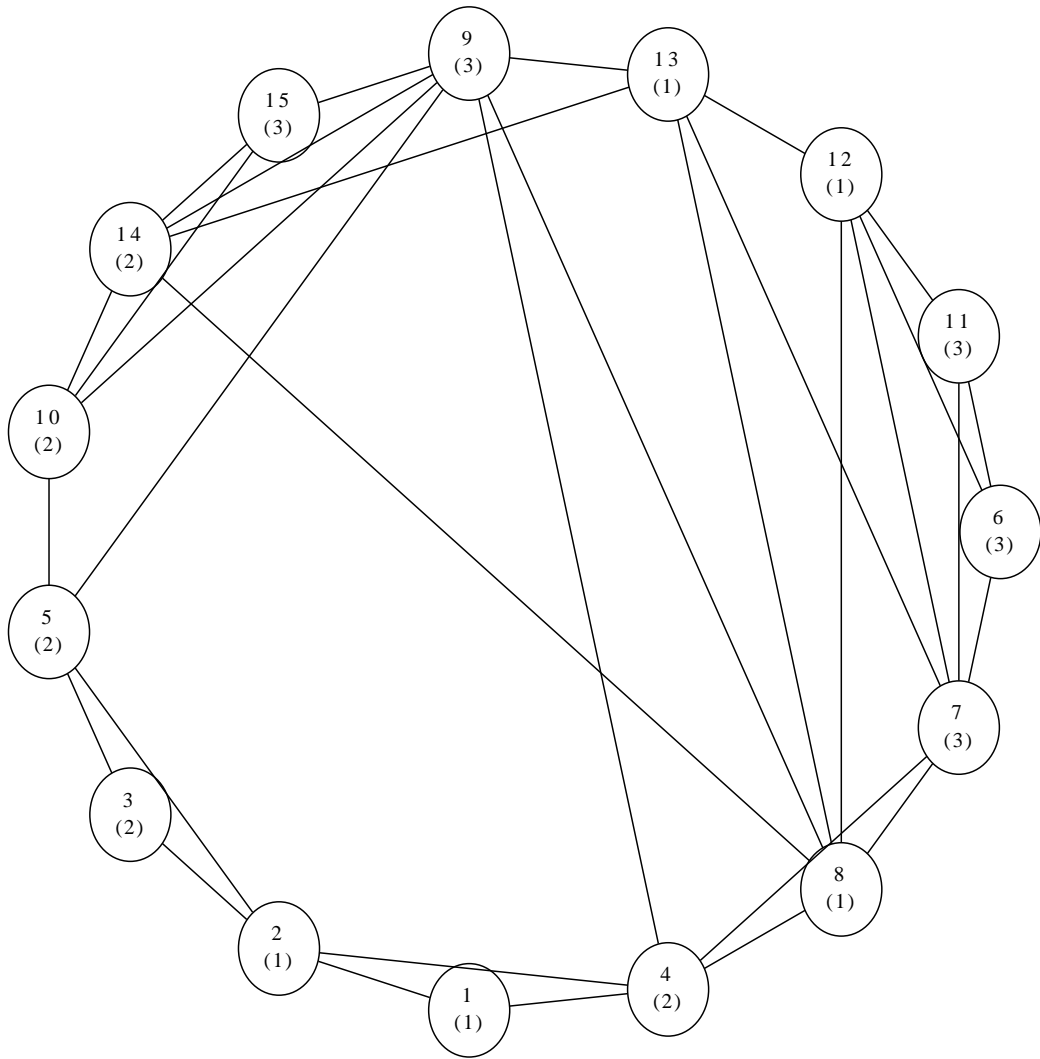


Figure 9: An experiment on an irregularly shaped field.

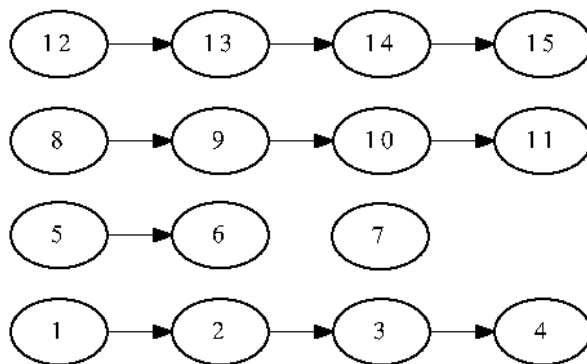


Figure 10: Cross-over trial with planned dropout. Nodes in network as subject/period combinations.

### 7.5 Example 7: Motivating example revisited

Recall that we assume we have a network of 10000 customers, and want to assess the effectiveness of sending 4 different adverts to a subset of 100 customers (before we send the best advertisement to the rest of the customers). An advertisement is deemed effective if more customers click on a hyperlink in a message. It can be determined which customers click on a hyperlink.

To generate a typical network of 10,000 customers; we start with two linked nodes and add 9998 further nodes individually according to a preferential attachment model, such that each node adds two links with probability of attachment to an existing node proportional to the degree of the existing node.

We then wish to assess designs with four adverts. We use five treatments: treatments 1 to 4 are coded to mean that adverts 1 to 4 are sent to a customer, and we invent a special treatment 0 which means that the customer is not selected. It is important that we include customers that are not given an advertisement in our experiment, as even though we do not show them an advertisement directly, the advertisement may be passed to them virally and they may click on the hyperlink.

As our network of  $n = 10000$  does not allow us to exhaustively search, we compare two methods: 1) the lowest information function found from a number of iterations of the exchange algorithm presented in section 6.3; 2) the lowest information function found after evaluating a number of randomly chosen designs.

The results are presented as Figure 11 by a plot of the lowest information function value found against the number of designs evaluated for each algorithm, on a log-log scale.

For both algorithms, the algorithm converges to find designs with low optimal function values. In this example, choosing random designs seems to work better than an exchange algorithm, although the difference in efficiencies between the two designs finally found (noting the log scale) is minimal.



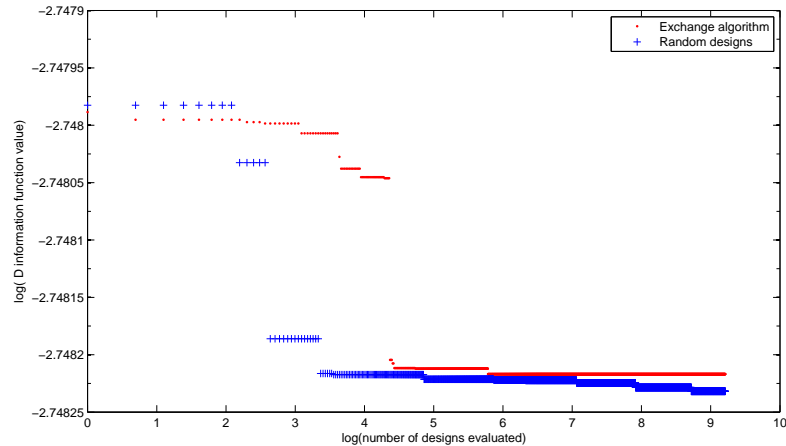


Figure 11: Comparison of algorithms for optimal design of large networks. Information function value against number of designs evaluated

## 8 Conclusions

We have presented a new methodology for designing experiments on social networks, an area which has seen little research. We introduced a linear network effects model that encompasses a wide class of experiments where the treatment on one subject affects the other subjects. We have shown how one can readily find optimal designs for these experiments, and described some techniques to make this exhaustive search procedure as fast as possible.

We have shown the problems in inflated bias and variance that may be experienced when we neglect a network effect in an experiment where one is present. Finally, we have presented examples from a wide area of experiments to demonstrate the flexibility of our method, focusing on marketing experiments on social networks, but also showing examples on agricultural and crossover trials. The method seems to be widely applicable and simple to implement, and we hope this paper serves as an introduction to a very wide class of models which can benefit from experimental design techniques.

## References

- J.-M. Azaïs and P. Druilhet. Optimality of neighbour balanced designs when neighbour effects are neglected. *Journal of Statistical Planning and Inference*, 64(2):353–367, 1997.
- R. Bailey. *Design of Comparative Experiments*. Cambridge University Press, 2008.
- J. Besag and R. Kempton. Statistical analysis of field experiments using neighbouring plots. *Biometrics*, 42(2):231–251, June 1986. ISSN 0006-341X. doi: 10.2307/2531047. URL <http://www.jstor.org/stable/2531047>.

- P. Doreian, L. Freeman, D. White, and A. Romney. Models of network effects on social actors. *Research Methods in Social Network Analysis*, pages 295–317, 1989.
- P. Druilhet. Optimality of neighbour balanced designs. *Journal of Statistical Planning and Inference*, 81(1):141–152, Oct. 1999. ISSN 0378-3758. doi: 16/S0378-3758(99)00004-X. URL <http://www.sciencedirect.com/science/article/pii/S037837589900004X>.
- A. Hedayat and W. Zhao. Optimal two-period repeated measurements designs. *The Annals of Statistics*, pages 1805–1816, 1990.
- B. Jones and M. G. Kenward. *Design and Analysis of Cross-Over Trials*. Chapman & Hall, 2003.
- J. Kunert and R. J. Martin. On the determination of optimal designs for an interference model. *The Annals of Statistics*, 28(6):1728–1742, Dec. 2000. ISSN 00905364. URL <http://www.jstor.org/stable/2674005>.
- J. Kunert and S. Mersmann. Optimal designs for an interference model. *Journal of Statistical Planning and Inference*, 141(4):1623–1632, Apr. 2011. ISSN 0378-3758. doi: 16/j.jspi.2010.11.020. URL <http://www.sciencedirect.com/science/article/pii/S0378375810005203>.
- R. Leenders. Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24(1):21–47, 2002.
- A. Lubiw. Some NP-complete problems similar to graph isomorphism. *SIAM Journal on Computing*, 10:11, 1981.
- M. McConnell, B. Sinclair, and D. P. Green. Detecting social networks: design and analysis of multilevel experiments. In *Third Annual Center for Experimental Social Science and New York University Experimental Political Science Conference*, 2010.
- S. C. Pearce. Experimenting with organisms as blocks. *Biometrika*, 44(1/2):141–149, June 1957. ISSN 0006-3444. doi: 10.2307/2333246. URL <http://www.jstor.org/stable/2333246>.
- D. B. Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6(1):34–58, 1978.